AMENDMENTS TO THE CLAIMS

1. (Currently Amended) A method for detecting speaker changes in an input audio stream comprising:

segmenting the input audio stream into predetermined length intervals;

decoding the intervals to produce a set of phones corresponding to each of the intervals;

generating a similarity measurement based on a first portion of the audio stream that is within one of the intervals and that occurs prior to a boundary between adjacent phones in one of the intervals and a second portion of the audio stream that is within the one of the intervals and that occurs after the boundary; ~~and~~

detecting speaker changes based on the similarity measurement; and

outputting an indication of the detected speaker changes.

2. (Original) The method of claim 1, wherein the predetermined length intervals are approximately thirty seconds in length.

3. (Original) The method of claim 1, wherein segmenting the input audio stream includes:

creating the predetermined length intervals such that portions of the intervals overlap one another.

4. (Currently Amended) The method of claim 1, wherein generating a similarity measurement includes:

calculating cepstral vectors for the audio stream prior to the boundary and ~~the audio stream~~ after the boundary, and

comparing the cepstral vectors.

5. (Original) The method of claim 4, wherein the cepstral vectors are compared using a generalized likelihood ratio test.

6. (Original) The method of claim 5, wherein a speaker change is detected when the generalized likelihood ratio test produces a value less than a preset threshold.

7. (Original) The method of claim 1, wherein the decoded set of phones is selected from a simplified corpus of phone classes.

8. (Original) The method of claim 7, wherein the simplified corpus of phone classes includes a phone class for vowels and nasals, a phone class for fricatives, and a phone class for obstruents.

9. (Original) The method of claim 8, wherein the simplified corpus of phone classes further includes a phone class for music, laughter, breath and lip-smack, and silence.

10. (Original) The method of claim 7, wherein the simplified corpus of phone classes includes approximately seven phone classes.

11. (Currently Amended) A device for detecting speaker changes in an audio signal, the device comprising:

a processor; and

a memory containing instructions that when executed by the processor cause the processor to:

segment the audio signal into predetermined length intervals,

decode the intervals to produce a set of phones corresponding to each of the intervals,

generate a similarity measurement based on a first portion of the audio signal that occurs prior to a boundary between phones in one of the sets of phones of an interval and a second portion of the audio signal that occurs after the boundary, ~~and~~

detect speaker changes based on the similarity measurement; and

store an indication of the detected speaker changes.

12. (Original) The device of claim 11, wherein the predetermined length intervals are approximately thirty seconds in length.

13. (Original) The device of claim 11, wherein segmenting the audio signal includes:

creating the predetermined length intervals such that portions of the intervals overlap one another.

14. (Original) The device of claim 11, wherein the set of phones is selected from a simplified corpus of phone classes.

15. (Original) The device of claim 14, wherein the simplified corpus of phone classes includes a phone class for vowels and nasals, a phone class for fricatives, and a phone class for obstruents.

16. (Original) The device of claim 15, wherein the simplified corpus of phone classes further includes a phone class for music, laughter, breath and lip-smack, and silence.

17. (Original) The device of claim 11, wherein the simplified corpus of phone classes includes approximately seven phone classes.

18. (Currently Amended) A device for detecting speaker changes in an audio signal, the device comprising:

a segmentation component configured to segment the audio signal into predetermined length intervals;

a phone classification decode component configured to decode the intervals to produce a set of phone classes corresponding to each of the intervals, a number of possible phone classes being approximately seven; and

a speaker change detection component configured to detect locations of speaker changes in the audio signal based on a similarity value calculated over a first portion of the audio signal that occurs prior to a boundary between phone classes in one of the intervals ~~sets of phone classes~~ and a second portion of the audio signal that occurs after the boundary in the one of the intervals, ~~sets of phone classes~~

wherein an indication of the detected locations of speaker changes are output from the device.

19. (Original) The device of claim 18, wherein the predetermined length intervals are approximately thirty seconds in length.

20. (Original) The device of claim 18, wherein the segmentation component segments the predetermined length intervals such that portions of the intervals overlap one another.

21. (Original) The device of claim 18, wherein the phone classes include a phone class for vowels and nasals, a phone class for fricatives, and a phone class for obstruents.

22. (Original) The device of claim 21, wherein the phone classes further include a phone class for music, laughter, breath and lip-smack, and silence.

23. (Currently Amended) A system comprising:

an indexer configured to receive input audio data and generate a rich transcription from the audio data, the rich transcription including metadata that defines speaker changes in the audio data, the indexer including:

a segmentation component configured to divide the audio data into overlapping segments of a predetermined length,

a speaker change detection component configured to detect locations of speaker changes in the audio data based on a similarity value calculated at locations in the segments that correspond to phone class boundaries;

a memory system for storing the rich transcription; and

a server configured to receive requests for documents and to respond to the requests by transmitting ones of the rich transcriptions that match the requests.

24. (Original) The system of claim 23, wherein the indexer further includes at least one of: a speaker clustering component, a speaker identification component, a name spotting component, and a topic classification component.

25. (Original) The system of claim 23, wherein the overlapping segments are segments of a predetermined length.

26. (Original) The system of claim 25, wherein the predetermined length is approximately thirty seconds.

27. (Original) The system of claim 23, wherein the phone classes include a phone class for vowels and nasals, a phone class for fricatives, and a phone class for obstruents.

28. (Original) The system of claim 27, wherein the phone classes additionally include a phone class for music, laughter, breath and lip-smack, and silence.

29. (Original) The system of claim 23, wherein the phone classes include approximately seven phone classes.

30. (Currently Amended) A device comprising:

means for segmenting the input audio stream into predetermined length intervals;

means for decoding the intervals to produce a set of phones corresponding to each of the intervals;

means for generating a similarity measurement based on audio within one of the intervals that is prior to a boundary between adjacent phones and based on audio within the one of the intervals that is after the boundary; ~~and~~

means for detecting speaker changes based on the similarity measurement; and

means for outputting the detected speaker changes.

31. (Original) The device of claim 30, wherein the predetermined length intervals overlap one another.